Title

Tri Nguyen

nguyetr9@oregonstate.edu

November 1, 2024

Given an unknown matrix $\boldsymbol{X} \in \mathbb{R}^{N \times N}$. We would like to learn this matrix given some signals about some of its elements. In particular, for some randomly chosen position (i, j), we obtain

$$y_{ij} \sim \text{some probability pdf with parameter } x_{ij}$$

There are 2 sources of difficulties:

- The set of indices $\Omega \subset [N] \times [N]$ usually does not cover the whole $[N] \times [N]$, leaving some positions in $\boldsymbol{X}$ never been implicitly observed.
- Even at the selected position (i, j), the observation $y_{i,j}$ is not exactly $x_{i,j}$ but only a random variable affected by $x_{i,j}$. This loss of information adds another layer of difficulty in recovering $\boldsymbol{X}$.

Luckily, the latter issue can be somewhat circumvented by acquiring more and more samples. The number of samples required for a certain accuracy depends on the pdf as well as the role of parameter x_{ij} to that pdf. Our interest is however, on the former issue.

For that, lets assume a pretty simple case:

$$\mathsf{Pr}\left(y_{ij} = 1\right) = x_{ij}.$$

This means that we use the Bernoulli distribution as the pdf, and x_{ij} is the mean of the distribution. Consider a simple unbiased mean estimator, i.e.,

$$\widehat{x}_{ij} = \frac{1}{|\Omega_{ij}|} \sum_{i=1}^{|\Omega_{ij}|} x_\omega,$$

then we know the sample complexity is

$$\mathbb{E}[(\widehat{x}_{ij} - x_{ij})^2] \leq$$

Now, lets unify the task. We wish to establish the following relation: Given an estimator $\widehat{\boldsymbol{X}}$ of $\boldsymbol{X}$ given observed dataset $\boldsymbol{S}$, the following holds with probability at least $1 - \delta$:

$$d(\widehat{\boldsymbol{X}}, \boldsymbol{X}) \leq g(S, \delta),$$

where d is some metric function, usually $\|\cdot\|_{\mathrm{F}}^2$ in this case, $g(S, \delta)$ is a non-decreasing function of number of samples and the probability $\delta \in [0, 1]$. In our goal specifically is to define d (easy part) and find the concrete function $g(S, \delta)$ so that the above statement holds true.

Why? Through several examples, we will see how the underlying structure of $\boldsymbol{X}$ affect sample complexity $g(S, \delta)$. In particular, we will go through following cases:

- No structure at all :)))
- Low rank structure, $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H}^\top$, where $\boldsymbol{W}, \boldsymbol{H} \in \mathbb{R}^{N \times K}, K \ll N$
- Neural network model: $x_{ij} = f(\boldsymbol{z}_{ij})$, where $\boldsymbol{z}_{ij} \in \mathbb{R}^d$ is certain obtained feature for position (i, j), and $f : \mathcal{D} \to [0, 1]$.

Each of these structures should be carefully selected and justified depending on application.